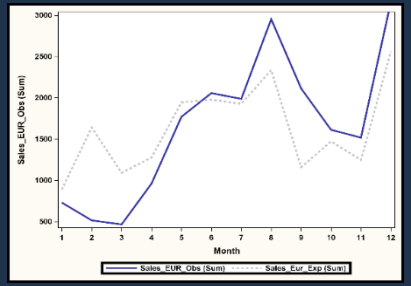
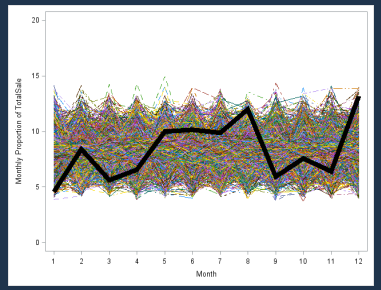




Data Science in Action #6

Checking the Alignment with Predefined Pattern

Gerhard Svolba
Data Scientist, SAS Austria

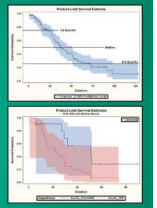
Data Science Applications and Case Studies

Data Science
in Action: #1

Performing Headcount Survival Analysis for Employee Retention

Can assumptions about the average length of time intervals be made, even if most of the endpoints have not yet been observed?

Survival analysis methods: Kaplan-Meier estimates
Cox Proportional Hazards regression
Survival Data Mining

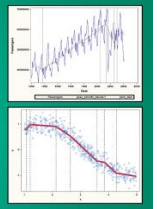


Data Science
in Action: #2

Detecting Structural Changes and Outliers in Longitudinal Data

Can events and changes in the course over time be automatically detected?

Smoothing Of Longitudinal Data
Multivariate Adaptive Regression Splines
Automatic Breakpoint Detection
Automatic Detection of Outliers with ARIMA Models

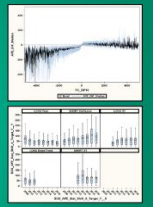


Data Science
in Action: #3

Explaining Forecast Errors and Deviations

Do the demand planners really improve forecast accuracy with their manual overwrites?

Linear Regression
Quantile Regression
Descriptive Statistics

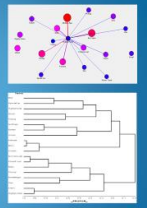


Data Science
in Action: #4

Listening to Your Data – Discover Relationships with Unsupervised Analysis Methods

Can your data tell you stories about your analysis subjects, even if you don't ask explicitly?

Unsupervised machine learning methods:
association analysis
variable clustering

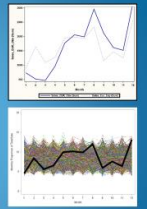


Data Science
in Action: #5

Checking the Alignment with Predefined Pattern

Which customers show a behavior that is far from what you expected?

Chi2 independency test
Benford's law
Time Series Similarity

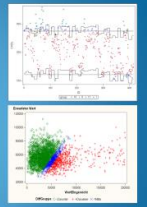


Data Science
in Action: #6

Proving a reference value that considers all available co-information

Can analytics help me to reduce the "Yes, but ... " sentences in my business discussions?

Linear Regression
Decision Trees
Time Series Analysis

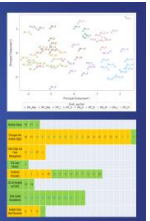


Data Science
in Action: #7

Topic Search Documents and Clustering

Can I automatically find clusters of documents with similar content?

Text Mining
Text Parsing (Synonyme, Stemming, Stop-Listen)
Term by Document Weights

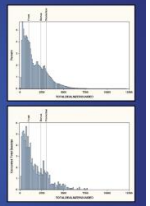


Data Science
in Action: #8

Using Monte Carlo Simulations to Understand the Outcome Distribution

When the sales manager looks at the project pipeline, does the sum of weighted averages give him or her a full picture?

Monte Carlo Simulations
Mathematical Programming

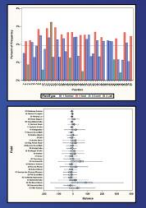


Data Science
in Action: #9

Studying Complex Systems – Simulating the Monopoly Board Game

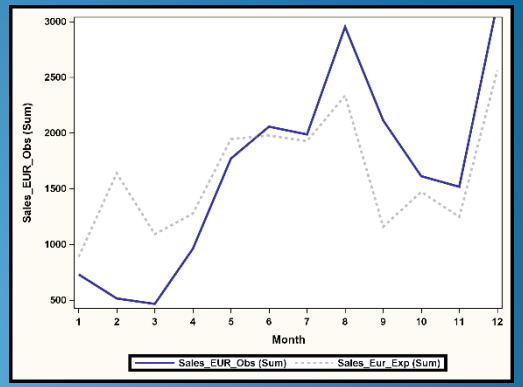
How can you simulate complex environments to get insight in the most frequent processes?

Monte Carlo Simulations

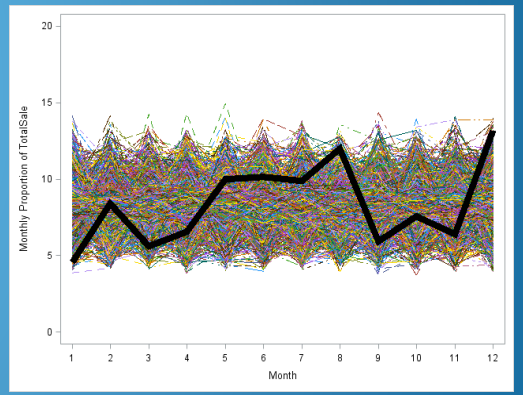


Data Science in Action: #6

Checking the Alignment with Predefined Pattern



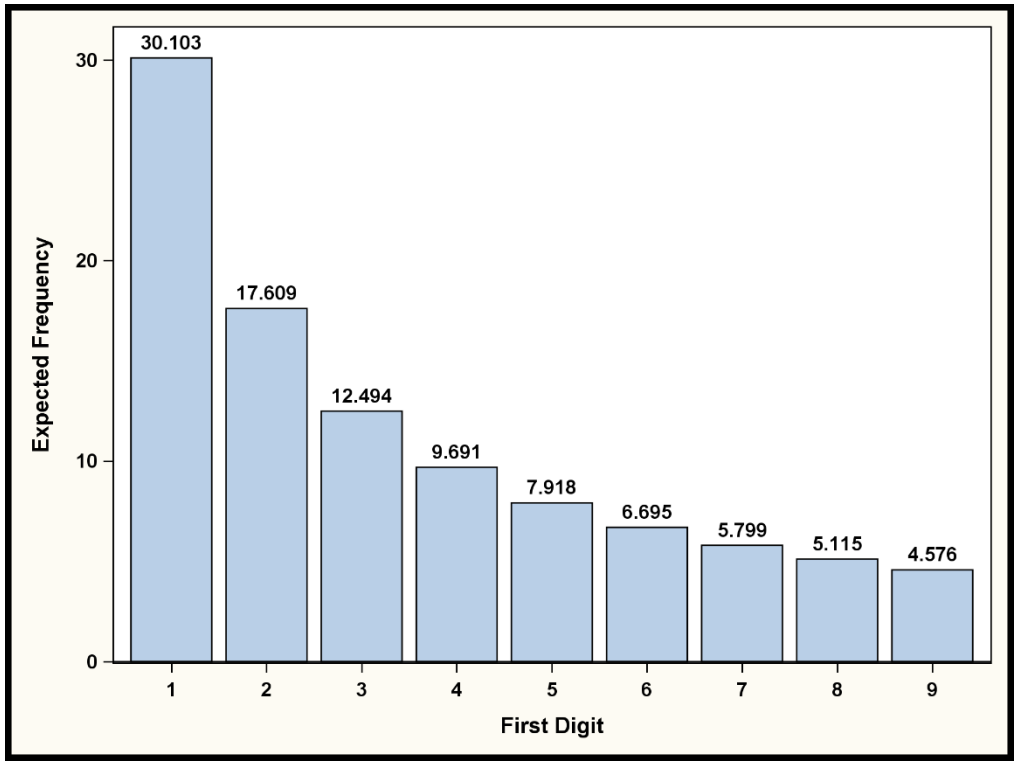
Which customers show a behavior that is far from what you expected?



Chi2 independency test
Benford's law
Time Series Similarity

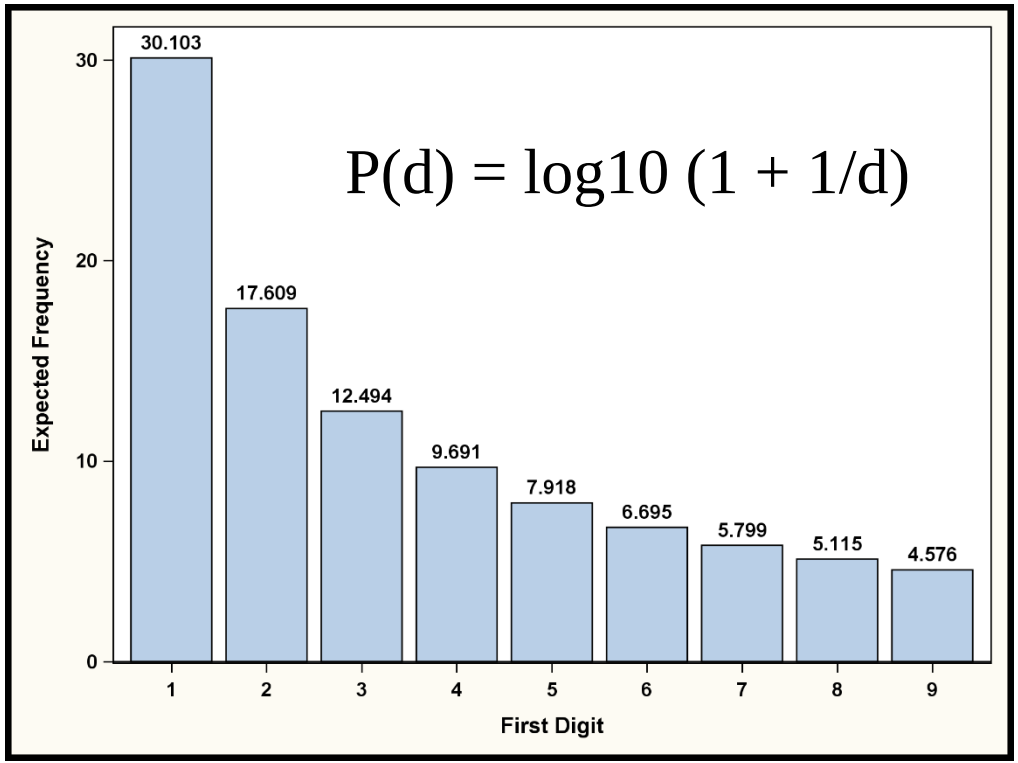
The Benford Distribution

Benford‘s Law – Distribution of the Digits 1-9




```
1,323.23
   43.00
  622.12
    1.10
   89.09
2,592.22
    7.40
   82.10
  620.19
   30.00
```

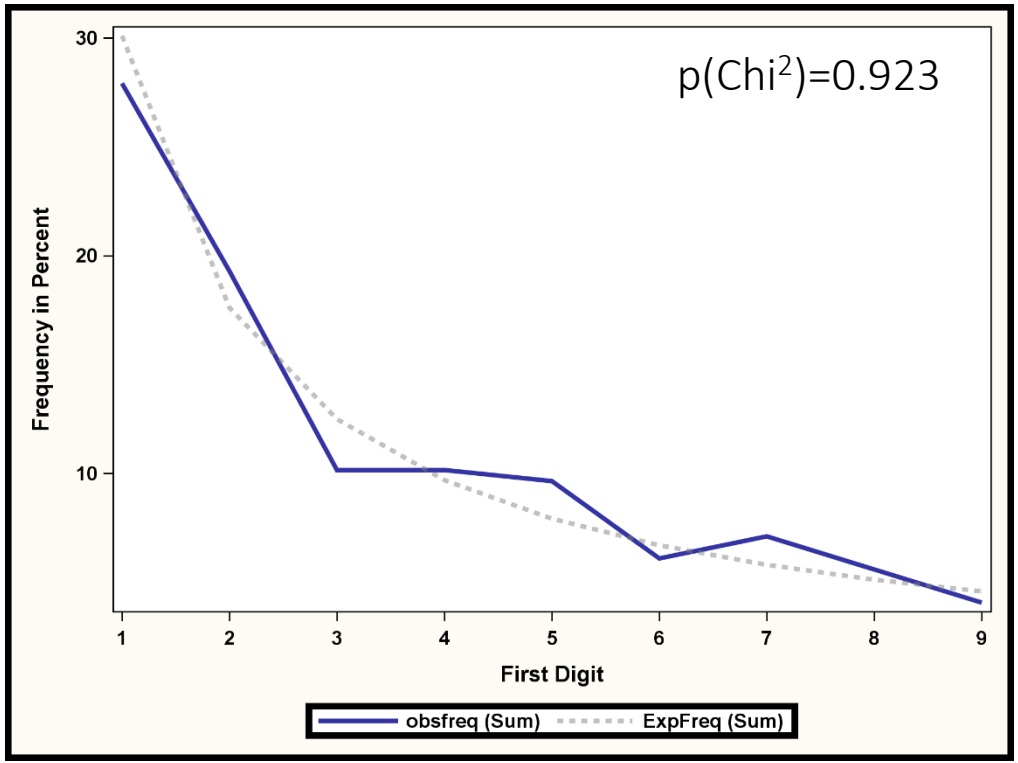
Benford‘s Law – Distribution of the Digits 1-9



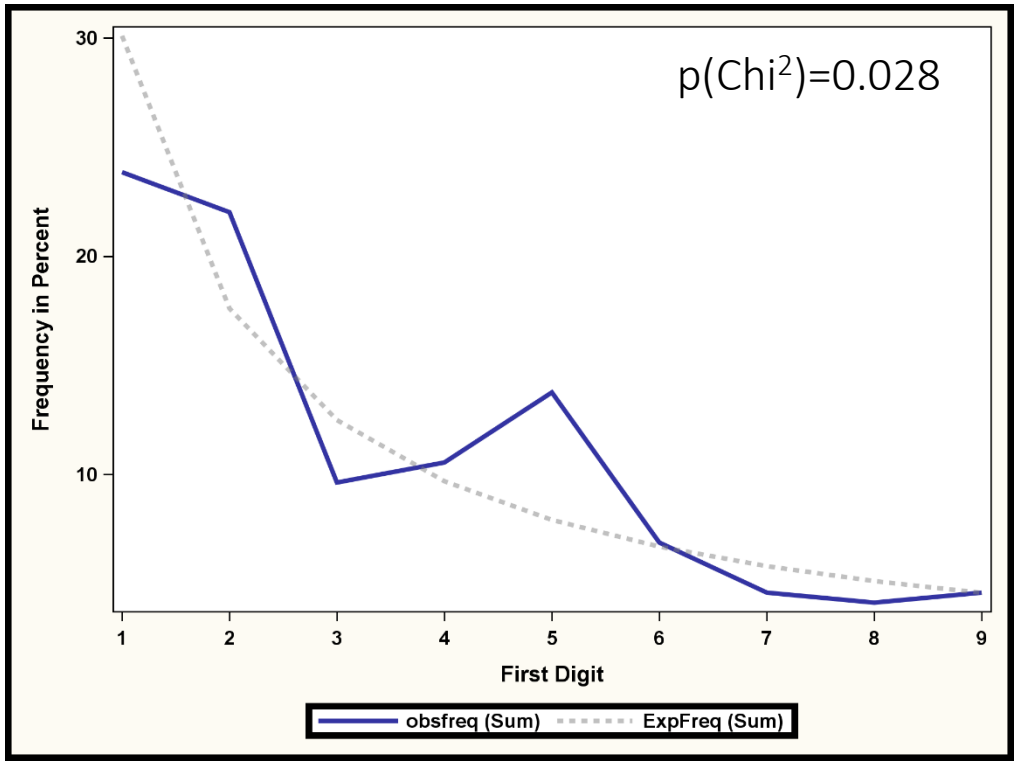
$$P(d) = \log10\ (1 + 1/d)$$

- 1881: Simon Newcomb
- 1938: Frank Benford
- 1972: Hal Varian

Distribution of Digits 1-9 vs Benford‘s Distribution for Account B



Distribution of Digits 1-9 vs Benford‘s Distribution for Account K



Rank the customer list by deviation from the expected distribution

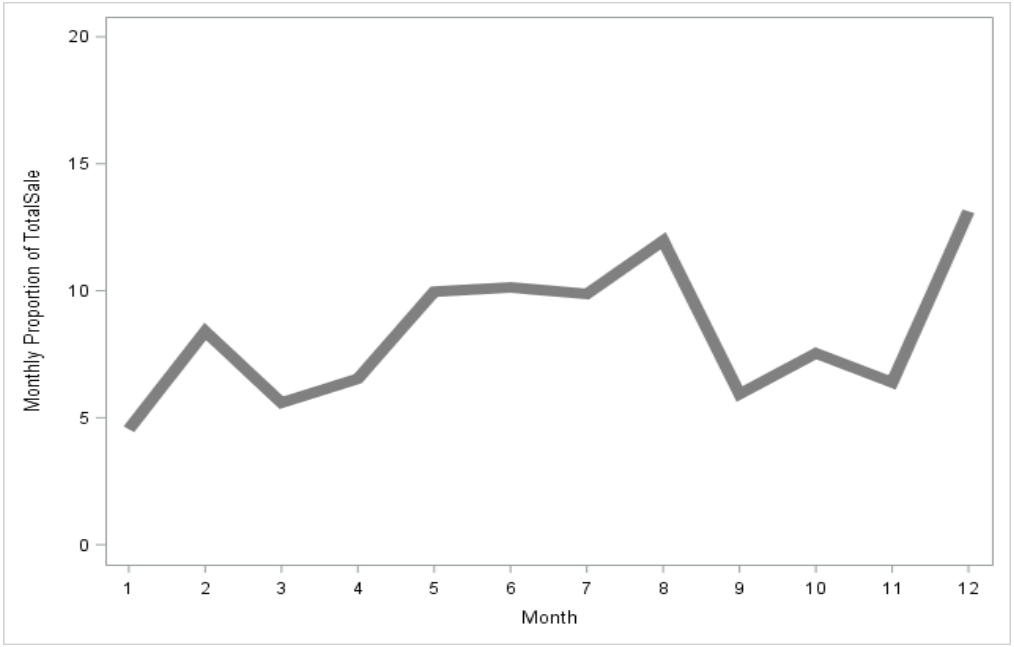
Rank	CustomerID	Chi2_Value	P_Value
1	5000	42.3	0.000%
2	2000	33.4	0.005%
3	8000	28.3	0.042%
4	4000	28.0	0.048%
5	3000	27.1	0.068%
6	1000	26.4	0.090%
7	10000	25.2	0.145%
8	6000	23.0	0.341%
9	11000	17.9	2.207%
10	7000	15.0	5.898%
11	9000	10.4	23.95%

Analysing the Accordance with pre-defined Pattern

Which of my sales representatives do not follow pre-defined pattern?

The demand for sub-contractors for a company in the catering business varies over the calendar year.

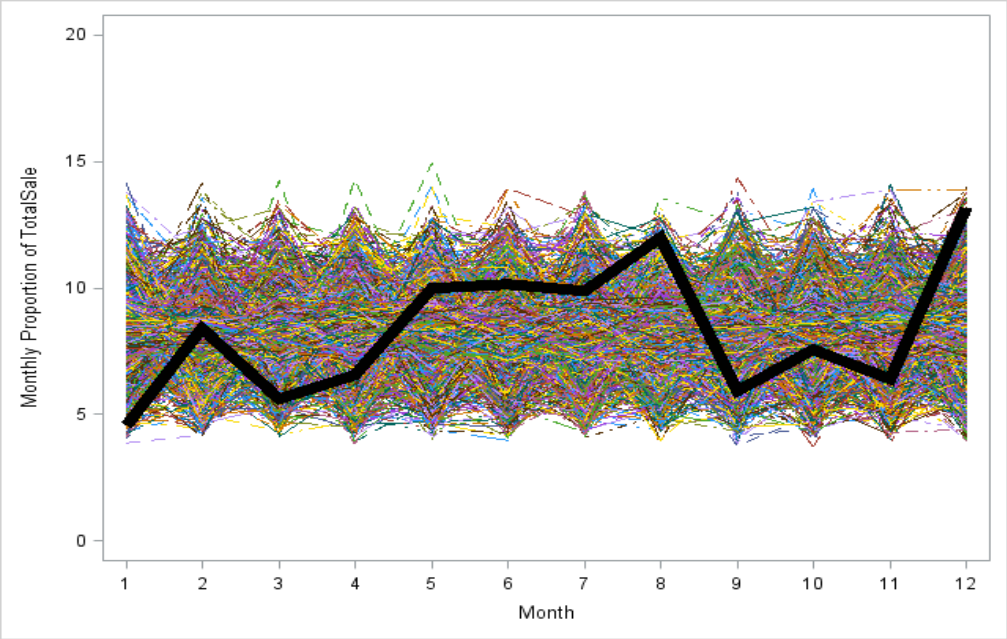
Sales Persons are forced to close such sub-contracts following the seasonal demand pattern.



Looking at the individual seasonal pattern per sales person does not help

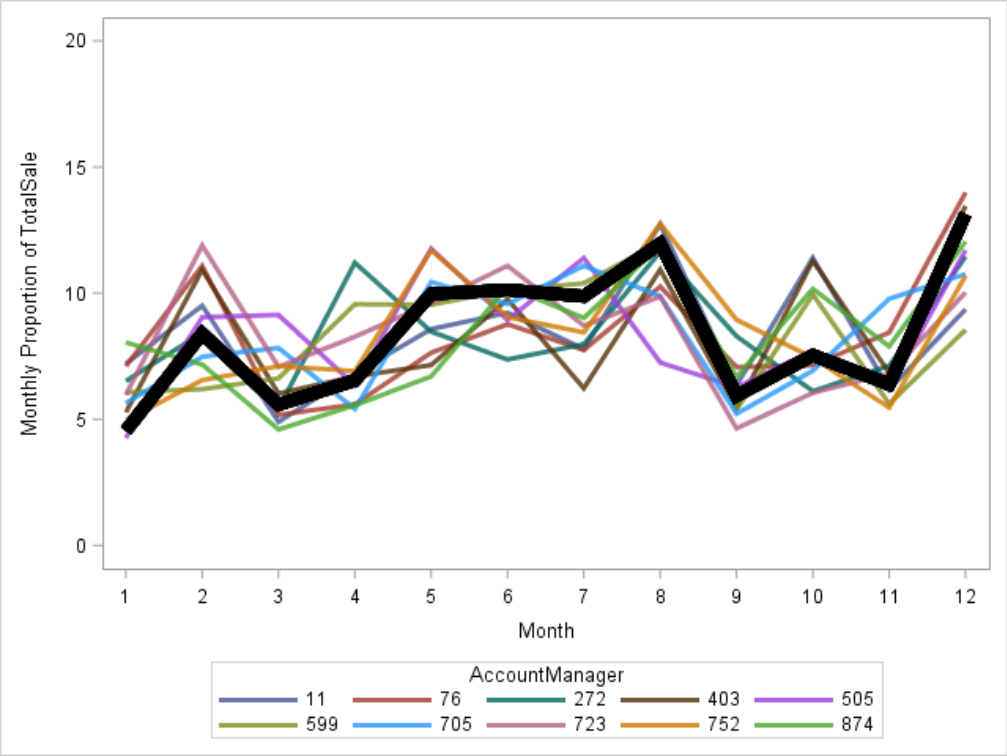
No clear picture.

Infeasible to review all individual lines manually.



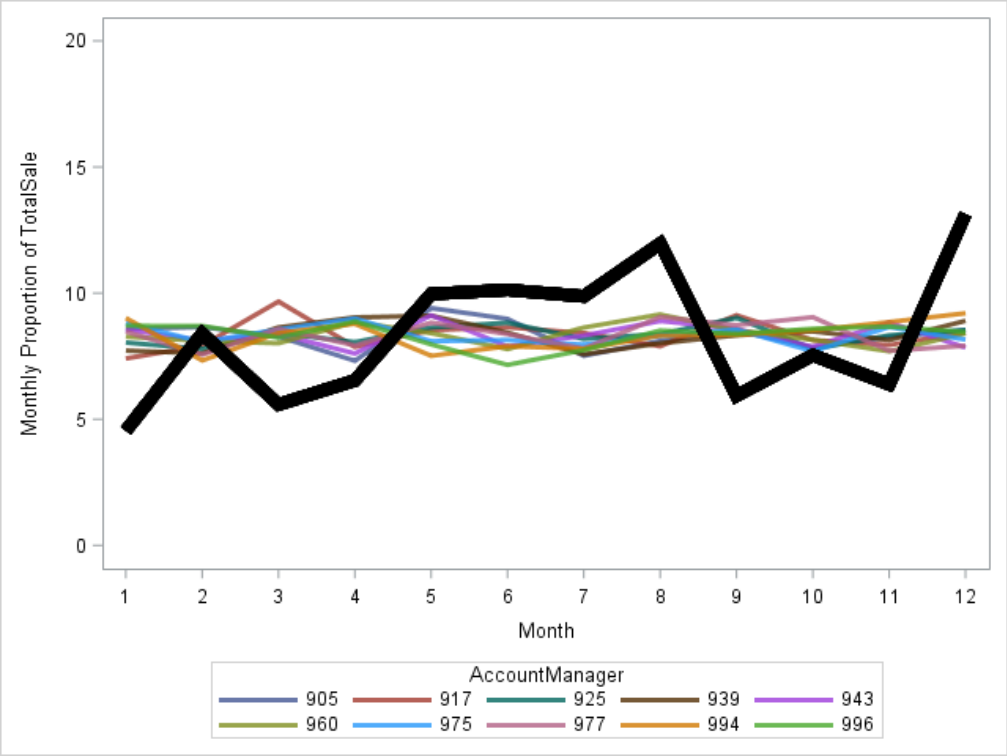
Use Analytical Methods to Rank Your Sales Persons (1)

Top 10 sales persons adhering to the pattern



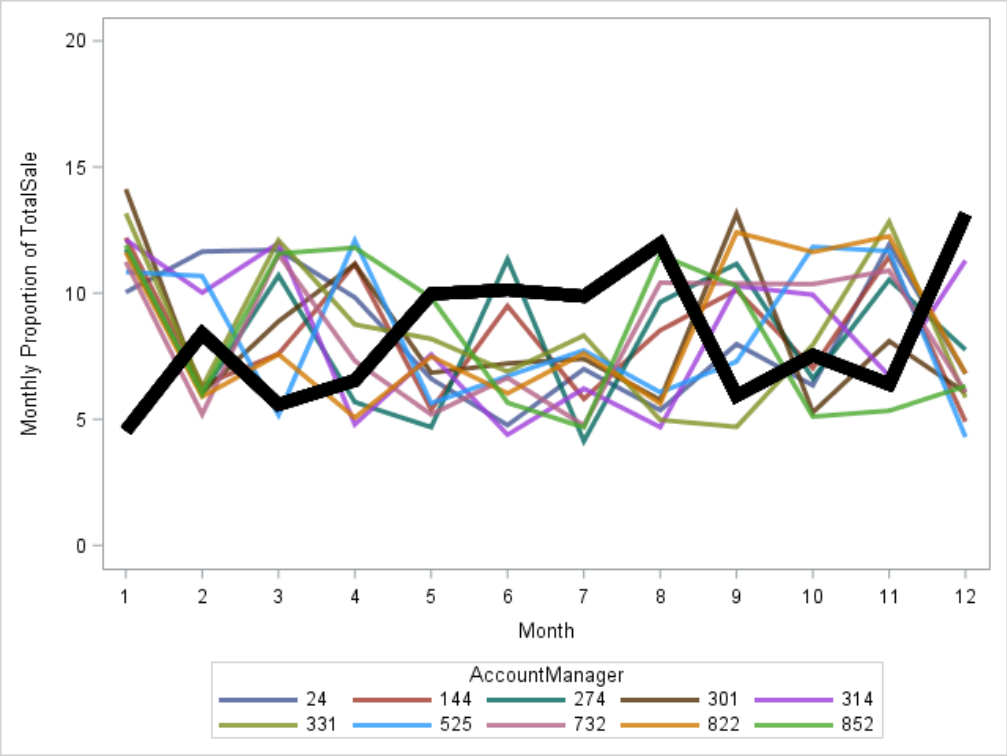
Use Analytical Methods to Rank Your Sales Persons (2)

10 sales persons without seasonal variation

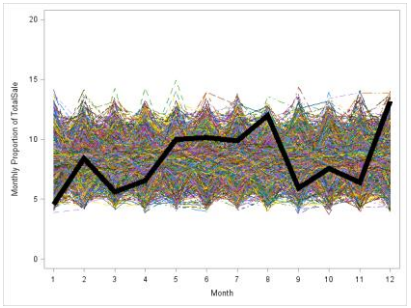


Use Analytical Methods to Rank Your Sales Persons (3)

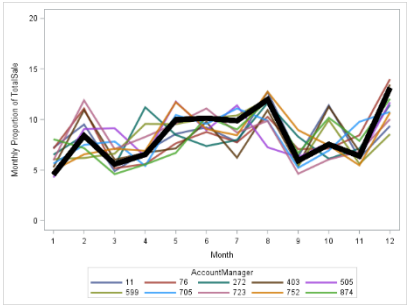
10 sales persons that work “against” the predefined pattern

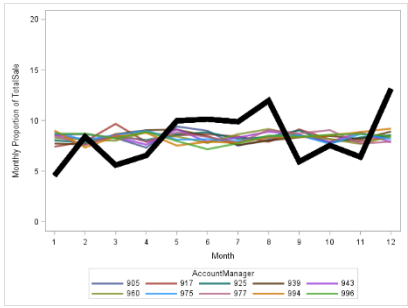


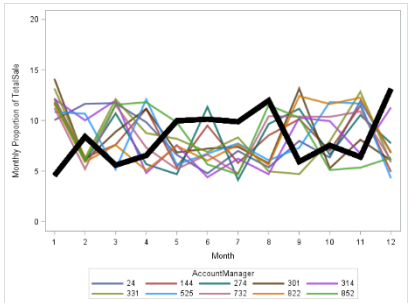
Analytics helps you, to get clearer picture!



Analytic Methods







From noise

to managable segments

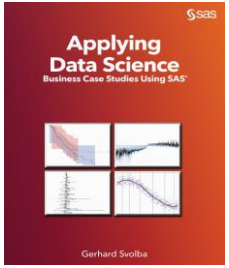
Conclusion

- Analytic methods allow to you quantify relations from the assumed distribution.
- Benford‘s law is often used in analysis of accounting data and in fraud analyics

- From noise to manageable segments - Analytics helps you, to get clearer picture!

Analytics and Data Science is there to help you!

- Get a clearer, more objective picture of your data and your analysis subjects
- Get explicit results instead of searching the needle in the haystack
- Make your data talk to you!
- Receive findings automatically instead of manually
- Do it again! – treat models as an asset and repeat your analysis



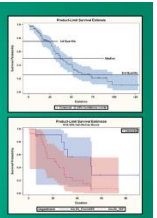
Data Science Applications and Case Studies

Data Science
in Action: #1

Performing Headcount Survival Analysis for Employee Retention

Can assumptions about the average length of time intervals be made, even if most of the endpoints have not yet been observed?

Survival analysis methods: Kaplan-Meier estimates
Cox Proportional Hazards regression
Survival Data Mining

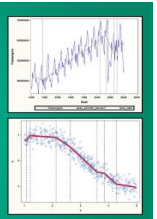


Data Science
in Action: #2

Detecting Structural Changes and Outliers in Longitudinal Data

Can events and changes in the course over time be automatically detected?

Smoothing Of Longitudinal Data
Multivariate Adaptive Regression Splines
Automatic Breakpoint Detection
Automatic Detection of Outliers with ARIMA Models

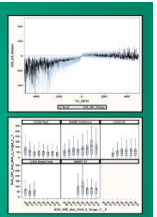


Data Science
in Action: #3

Explaining Forecast Errors and Deviations

Do the demand planners really improve forecast accuracy with their manual overwrites?

Linear Regression
Quantile Regression
Descriptive Statistics

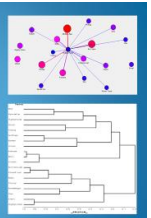


Data Science
in Action: #4

Listening to Your Data – Discover Relationships with Unsupervised Analysis Methods

Can your data tell you stories about your analysis subjects, even if you don't ask explicitly?

Unsupervised machine learning methods:
association analysis
variable clustering

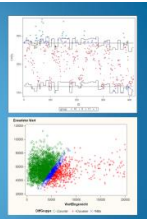


Data Science
in Action: #6

Proving a reference value that considers all available co-information

Can analytics help me to reduce the "Yes, but ... " sentences in my business discussions?

Linear Regression
Decision Trees
Time Series Analysis

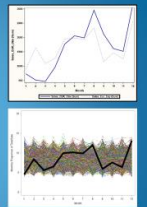


Data Science
in Action: #5

Checking the Alignment with Predefined Pattern

Which customers show a behavior that is far from what you expected?

Chi2 independency test
Benford's law
Time Series Similarity

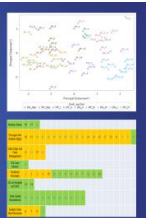


Data Science
in Action: #7

Topic Search Documents and Clustering

Can I automatically find clusters of documents with similar content?

Text Mining
Text Parsing (Synonyme, Stemming, Stop-Listen)
Term by Document Weights

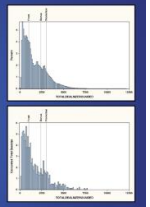


Data Science
in Action: #8

Using Monte Carlo Simulations to Understand the Outcome Distribution

When the sales manager looks at the project pipeline, does the sum of weighted averages give him or her a full picture?

Monte Carlo Simulations
Mathematical Programming

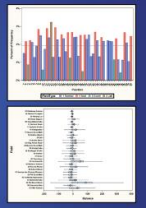


Data Science
in Action: #9

Studying Complex Systems – Simulating the Monopoly Board Game

How can you simulate complex environments to get insight in the most frequent processes?

Monte Carlo Simulations



Get access to more content:

SAS DACH @Youtube:	https://www.youtube.com/user/SASsoftwareGermany
Blogs on LinkedIn:	https://www.linkedin.com/in/gerhardsvolba/
Twitter:	https://twitter.com/gsvolba
Content on Github:	https://github.com/gerhard1050
Books @SAS-Press:	https://support.sas.com/svolba